



401 Bruce Hall
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PENNSYLVANIA 15260

MACHINE INTELLIGENCE SYSTEMS LABORATORY

October 28, 1974

RECEIVED
OCT 30 1974
BFB-DNR

Dr. Carl R. Brewer
Biotechnology Resources Branch
Division of Research Resources
Department of Health, Education and Welfare
Public Health Service
National Institutes of Health
Bethesda, Maryland 20014

OCT 30 1974
BFB-DNR

Dear Dr. Brewer:

Our laboratory was recently designated one of a small number of Computer Laboratory Health Care Resources to be funded under provisions of the Computer Manpower Training Act of 1971 (Section 769 a(1), P.L. 92-157). This resource grant, which is administered by the Bureau of Health Resources Development (BHRD), has provided for support of personnel, data terminals, modest communications costs, and other incidental expenses that will be incurred during the next three years in connection with our Medical DIALOG project. The grant does not, however, provide for access to suitable computer facilities to support this work; instead, we have been directed to address our computational needs to the SUMEX/AIM resource.

Briefly, the DIALOG (for DIAGnostic LOGic) project is a large scale computer-based medical diagnosis system that makes use of the methods and structures of artificial intelligence. Unlike most other computer diagnostic programs, which are oriented to the differential diagnosis of disease in narrowly circumscribed areas, the DIALOG system has been designed to deal with the general problem of diagnosis in internal medicine. The approach taken has been one of cognitive simulation, with data structures and programs devised to mimic those of the expert diagnostician. A medical data base has been developed, which at the present time encompasses perhaps twenty-five percent of the major diseases of internal medicine. Initial tests of the DIALOG procedure in the diagnosis of illnesses that fall within the domain of this data base have shown a high rate of success, even in difficult clinical problems complicated by the concurrence of several distinct disease entities.

Our intention in seeking the BHRD grant was to secure support for extensive expansion of the data base and concurrent evaluation and refinement of the diagnostic procedure. In addition, our plans call for the adaptation of the system to subserve educational as well as diagnostic purposes. As presently envisioned, a higher-level diagnostic process would be invoked to manipulate the underlying DIALOG models so as to infer, evaluate, and criticize the diagnostic process exhibited by students (or others) engaged in interactive problem solving with the system. Such a meta-diagnostic process would have obvious value in teaching; perhaps equally important applications might be found in professional evaluation and review.

In order to achieve these objectives, we must have access to computer facilities capable of handling the special requirements of large scale artificial intelligence applications. In view of the fact that the SUMEX/AIM facility has been designed specifically to support Artificial Intelligence in Medicine, and because our objectives are in such close consonance with those of the AIM program, we hope to secure the requisite level of computational support via this national shared resource.

Our requirements cannot be stated precisely at this point because the operating system and language processors available on SUMEX/AIM differ significantly from those currently being used. The present version of DIALOG runs on a DECsystem-10 facility comprising a KI10 processor and 196k of installed core; standard software includes the DEC 1070 monitor and LISP 1.6 language processors.

As presently implemented in LISP 1.6, the DIALOG system requires 135k words of main memory, apportioned approximately as follows:

LISP interpreter	5k
DIALOG programs (compiled)	15k
Data Base	90k
Working Storage	15k
Misc. (stacks, bit tables, etc)	10k

Utilization of other system resources includes:

*For each clinical problem analyzed, from one to three minutes of cpu time

*For permanent storage of programs and data, approximately one million words of disk space

Although the present system, complete with its own LISP interpreter, could be brought up under TENEX and operated in approximately one half of the available 262k/user address space, it would be clearly undesirable to make extensive use of a LISP interpreter not specifically designed for use in the demand paging environment of TENEX. We have therefore decided to undertake the necessary changes to convert DIALOG from LISP 1.6 to INTERLISP.

We anticipate some problems with this conversion, since the INTERLISP processor itself is understood to occupy something in excess of 160k words of the user address space. While existing programs and data might be compacted to fit within the remaining 100k of user core, it seems clear that no significant expansion can take place without major changes in the systems being used. One development that we hope to exploit is the effective extension of user address space made possible by use of overlay techniques in the 'shadow' version of INTERLISP now being readied. Similar techniques for mapping of data as well as program segments will probably be required in order to accommodate the eventual four-fold expansion of the data base which is contemplated.

Uncertainties as to the impact of these memory management strategies complicates the task of estimating resource requirements for support of DIALOG on SUMEX/AIM. However, it does seem clear that from the outset and throughout the life of the project, the DIALOG system will be pushing the limit of available user core.

During the first three to six months of the project, our efforts will be largely directed toward the conversion of programs and data. We estimate the following levels of resource utilization during this period:

No. of users active at one time	3 - 5
Disk space required	2 million words
CPU utilization	one hour per day
Terminal connect time	8 - 10 hours per day


Once the conversion process is completed, data base accretion and the planned research on meta-diagnostic processes will continue to generate additional demands: thus, a base level of research and development activity comparable to that of the conversion period may be expected to continue indefinitely.

In addition, the evaluation process will generate requirements that will be largely a function of the number of diagnostic studies analyzed daily. When field testing commences locally (targetted for Spring 1975) we estimate that the case load will grow from an initial rate of four or five per day to a level of perhaps twenty per day when fully operational at this site. By January, 1976, we expect additional field test sites to be brought on line, with as many as six such remote locations eventually becoming involved in the field test program: these will give rise to a potential maximal load of 120 to 140 diagnostic studies per day. Based on present averages (two minutes of cpu time and 30 minutes of connect time per case) this could mean an eventual load of four to five hours of cpu time and sixty to seventy hours of terminal connect time per day.

These are crude estimates, of course, based on incomplete information and constrained only by our determination for orderly implementation and evaluation of the system. As our experience with the SUMEX/AIM resource and its community of users grows, the estimate of our requirements will hopefully become both more precise and more realistic.

We trust that this brief statement has conveyed something of the significance of the DIALOG project, and some insights as to our intentions and apprehensions as we approach the transition to SUMEX/AIM. Despite the inevitability of conversion/start up costs, we feel that the move will be of great benefit and may in fact be critical to the ultimate success of the project. We appreciate your consideration of this request, and will be very pleased to supply any further information that may be required.

Sincerely yours,


Harry E. Pople, Ph.D.


Jack D. Myers, M.D.

mj

cc: Dr. Joshua Lederberg